

Does Training in Table Creation Enhance Table Interpretation? A Quasi-Experimental Study
With Follow-Up

Bryan T. Karazsia and Kendal Wong

The College of Wooster

Author Note

Bryan T. Karazsia, Department of Psychology, The College of Wooster; Kendal Wong, Department of Psychology, The College of Wooster.

Kendal Wong is now at the Department of Psychology, University of Chicago at Illinois.

This research was supported by a Scholarship of Teaching and Learning Grant from the Society for the Teaching of Psychology. The author also thanks Madelaine Flynn (The College of Wooster) for her assistance with coding.

Correspondence concerning this article should be addressed to Bryan T. Karazsia, Department of Psychology, The College of Wooster, Wooster, OH 44691. E-mail: bkarazsia@wooster.edu

Abstract

Quantitative and statistical literacy are core domains in the undergraduate psychology curriculum. An important component of such literacy includes interpretation of visual aids, such as tables containing results from statistical analyses. This paper presents results of a quasi-experimental study with longitudinal follow-up that tested the effectiveness of a new technique for enhancing student interpretation of APA-formatted tables. Undergraduate students exposed to the technique performed better than students not exposed on measures of table interpretation. Effect sizes between groups were large, even after a 3-month follow-up assessment. An active learning experience in which students learn how to create tables can enhance student ability to interpret tables presented in empirical psychological literature.

Keywords: table interpretation, statistics, quantitative literacy, statistical literacy, APA format

Does Training in Table Creation Enhance Table Interpretation? An Experimental Study With Follow-Up

Quantitative methodology is “a unifying discipline” in psychology (Aiken, West, & Millsap, 2008, p. 32). At the undergraduate level, quantitative and statistical literacy are core learning domains (e.g., Dunn et al., 2010), of which one important component includes accurate interpretation of visual aids (American Psychological Association [APA], 2007). Recently, a new learning technique designed to enhance student ability to interpret tables by training students how to create tables was developed and evaluated empirically (Karazsia, 2013). The purpose of the present study was to evaluate this method with a more rigorous design with longitudinal follow-up.

In an effort to enhance the table-interpretation ability of undergraduate students, Karazsia (2013) introduced a new learning technique to his students. The idea for this technique stemmed from observations of student difficulty interpreting tables in advanced psychology courses where reading empirical literature is required. The students were not transferring knowledge learned in an introductory statistics course to their advanced courses. Adopting terms from the cognitive sciences, students were demonstrating outcomes associated with shallow learning (e.g., Aleven & Koedinger, 2002) or shallow processing (Craik & Tulving, 1975). Karazsia designed the technique to enhance student engagement with course material. Instead of focusing on table interpretation explicitly, the new technique promotes active student engagement with statistics by helping students learn how to create APA-formatted tables to present results of analyses they conducted. This process involves understanding what statistics to use and where to place them in a table (e.g., corresponding with row and column headings). Karazsia hypothesized that this activity would promote greater involvement with material (e.g., focusing on the meaning of the

statistics and the tables, instead of only focusing on the numbers).

As reported by Karazsia (2013), students created their own unique APA-formatted tables as a homework assignment. Students exposed to the new learning technique performed better on an assessment of table interpretation ability than students not exposed to the technique.

Although the study revealed that the new teaching technique had promise, the investigation was limited in that the control group was one of convenience, and there was no longitudinal follow-up. Therefore, any demonstrated difference between the experimental and control groups could have been due to factors other than experimental exposure (such as recency of explicit statistical instruction). Additionally, even if the technique promoted students' learning in the short term, how long the gains persisted remained unknown.

We designed the present study to offer a more rigorous investigation of the table creation teaching technique. In this study, both the control and experimental groups consisted of students enrolled in one of two sections of an introduction to statistics course for behavioral sciences, which is a prerequisite to all advanced courses in the psychology major at the investigators' home institution. In the first assessment, researchers exposed one group of students to the new learning technique (experimental group); the second group was not exposed (control group). For the second assessment, researchers exposed both groups to the new learning technique. The decision to expose both groups to the technique was made to rule out the possibility that confounding factors, due to a quasi-experimental design instead of using random assignment, may explain group differences. We hypothesized that students in the experimental group would perform better than students in the control group for the first assessment (Hypothesis 1) and that there would be no difference in performance between groups for the second assessment

(Hypothesis 2). We expected these similarities and differences between group performances to remain during a 3-month follow-up assessment of student learning (Hypotheses 3 and 4).

Method

Participants

A total of 48 undergraduate students enrolled in one of two sections of an introductory statistics course for behavioral sciences participated in this study. Students in both sections were taking this course as a prerequisite for advanced topics in psychology or a related discipline. The mean age of the study sample was 18.83 years ($SD = 0.81$), with the majority of the sample self-reporting biological sex as female (64.58%). Racial composition of the sample was as follows: 75% White, 8.33% African American, 6.25% Asian, and 10.42% other (including multiracial). There were also four participants (8.33%) who self-reported ethnicity as Hispanic, Latino, or Hispanic origin.

Materials

Table interpretation assessment. We used four assessments to evaluate how students interpreted APA-formatted tables. The two primary assessments from the initial evaluation (see Karazsia, 2013) assessed student interpretation skills of a table presenting demographic information and a correlation matrix. We developed two new assessments for the longitudinal follow-up. They relied on examples from published literature that presented demographic information in a table (Karazsia, Guilfoyle, & Wildman, 2012) and a correlation matrix (Karazsia, Crowther, & Galioto, 2013).

All table assessments followed the same format. We presented students with a table followed by six multiple-choice and open-ended questions. For the multiple-choice questions, we instructed students to select the item representing the best possible answer. Several general

questions applied to all tables. For example, the item “Results from what statistical analyses are displayed in this Table?” had the following response choices: descriptive statistics, *t* tests, correlations, analyses of variance, and regressions. Another general item, “For (specific variable in respective table), what is the mean score reported in the Table?”, was open ended. As demonstrated in this example, even the opened-ended questions had concrete answers, which facilitated scoring. A third sample item, specific to the initial correlation table assessment, “Is ethnic identity associated with suicidality?” included the following response options: yes, no, it depends, and no application. This multiple-choice question was followed by open-ended prompt asking students to explain their previous answer. The scoring rubric specified that the correlation matrix revealed significant correlations between ethnic identity and suicidality among African American participants but not European American participants (see table as presented in APA, 2010, p. 136). The complete assessment instruments are available from the authors upon request.

Math ability. A 15-item, multiple-choice measure used in previous research (Johnson & Kuennen, 2006) was adopted to assess basic algebraic skills. We assessed this construct because mathematical ability is a predictor of student achievement in introductory statistics courses (Johnson & Kuenne, 2006; Lunsford & Poplin, 2011). Across the 15 items scored as correct versus incorrect, the internal consistency in the present sample was adequate (Cronbach’s $\alpha = .76$).

Procedure

A human participant review panel approved all procedures associated with this study. We randomly assigned the sections to be in either the control or experimental group. Prior to the table creation exercises, students in both the control and experimental conditions were provided

in-class instruction on the respective statistical techniques and computer-based instruction on how to analyze data (e.g., computing scale scores, running descriptive statistics, analyzing correlations). For the psychometric properties table, students in the experimental condition created an APA-formatted table that presented means, standard deviations, sample size, ranges, and distributional properties of scale variables. For the correlation table, students in both the experimental and control groups created an APA table that presented means, standard deviations, and intercorrelations between variables. To facilitate instruction on table creation, the professor of the course made updated templates from an online resource (Acock, van Dulmen, Kurdeck, Buehler, & Goldscheider, n.d.) available to all students. The minor updates ensured that tables conformed to the 6th edition of the APA (2010) *Publication Manual*.

The professor instructed students in both groups to turn in a unique assignment within 1 week, though collaboration was permissible. If students needed assistance, they were encouraged to meet with the professor. Within 2 days after the students turned in the assignments, the professor reviewed each assignment and provided feedback. The feedback consisted of handwritten corrections about errors in APA formatting or inaccuracies regarding the data reported. The professor graded assignments as part of standard procedures for the course (i.e., homework points that count toward the final grade). Two weeks after the assignment was due, approximately 12 days after students received graded feedback on their tables, a teaching assistant with no knowledge of study hypotheses administered the respective assessment of table interpretation skills. The professor left the room while the teaching assistant administered these assessments. Students had 10 min to read the consent form and complete the assessment. Completion of the assessments was anonymous, voluntary, and not incorporated in grading for the course.

The students in the control condition completed the same assessment for descriptive statistics on the same day and in the same manner as those in the experimental condition. For the second assessment on correlation tables, the professor instructed both the experimental and control conditions to construct a matrix table as a homework assignment using the same aforementioned procedure. Both groups then completed the same assessment on correlation tables using the same procedure as the first assessment.

After the professor submitted final grades for each class, a coder with no knowledge of study hypotheses or group designation scored the assessments using a detailed rubric. Scoring procedures yielded a composite score for each assessment, ranging from 0–6, with higher scores indicating higher comprehension of table interpretation.

Three months after the semester was finished and before the next semester began (August 2013), the professor invited all students from both classes to participate in a follow-up online assessment that contained two assessments of table interpretation skills: one of descriptive statistics and one of a correlation matrix.

Results

Prior to conducting the primary analyses, we imputed missing data from baseline assessments (0.51% of datapoints) using a Hotdeck imputation procedure (Myers, 2011). We did not impute missing data due to attrition during the follow-up assessment because patterns of missingness were not random. All continuous variables at both timepoints were sufficiently normally distributed (Tabachnick & Fidell, 2013). We compared baseline characteristics of the samples (ethnicity, age, math ability), and no significant differences emerged for either the total sample or subsample available for follow-up analyses (see Table 1).

1. We used independent-samples t tests to test all four study hypotheses (see Table 1). Hypothesis 1 stated that students in the experimental condition would perform better than those in the control condition on the demographic properties table assessment, and this hypothesis was supported ($d = 0.68$, $CI_{95\%}: 0.08\text{--}1.27$). Hypothesis 2 stated that there would be no significant difference between the two groups on the second assessment because we exposed both groups to the novel teaching technique, and this hypothesis was supported. Hypothesis 3 stated that, at the 3-month longitudinal follow-up, participants in the experimental condition would perform better than students in the control condition on a demographic properties table assessment. Results of this independent-samples t test were not statistically significant, although there was a statistical trend in the predicted direction ($p = .09$), and the effect size remained large according to Cohen's (1992) standards ($d = 0.77$; $CI_{95\%}: -0.11\text{--}1.59$). It is worth noting that (a) our use of a non-directional (two-tailed) test is conservative given that we specified a direction of results a priori, and (b) results were statistically significant with a directional (one-tailed) test ($p = .04$). Thus, results partially supported this hypothesis with two-tailed tests, yet they supported this hypothesis unequivocally with a one-tailed test. Hypothesis 4 stated that the two conditions would not differ on an assessment of correlation matrix interpretation skills during the 3-month follow-up, and results this hypothesis because the groups did not differ on this assessment during follow-up.

To expand on these analyses, we used a meta-analytic technique described by Rosenthal and Rosnow (2008) to compare the effect sizes obtained in the present study with those from the

original Karazsia (2013) study. As described by Rosenthal and Rosnow, two effect sizes can be compared via a z test after they are transformed into Fischer z scores. Only effect sizes corresponding to Hypotheses 1 and 3 of the present study were comparable to the effect size that Karazsia reported for differences between groups on the assessment of descriptive statistics interpretations. Karazsia reported an effect size of $d = 0.83$. Obtained z scores for the comparison between results in the present study with those reported previously were $z = 0.31$ and $z = 0.08$ (Hypotheses 1 and 3, respectively). Neither z score was statistically significant, indicating that the effect sizes obtained in the present study are comparable to the effects obtained in previous research.

Discussion

This study was the first to test the novel teaching technique proposed by Karazsia (2013) with a quasi-experimental design with longitudinal follow-up, and the results supported those presented by Karazsia. When students in the experimental group created an APA-formatted table, which the students in the control group did not create, they demonstrated better table interpretation skills. When students in both groups participated in the table creation exercise, there were no differences between groups on table interpretation abilities. It is important to note that we did not ask the students who created the APA tables to interpret APA tables prior to completing the assessment. Instead, students received instruction on how to create the tables using data collected from class for demonstrative purposes. Therefore, students demonstrated a transfer of knowledge from one task (table creation) to another (table interpretation), highlighting the effectiveness of an active learning technique.

Unlike the previous study by Karazsia (2013), there were no significant differences between the control and experimental groups at baseline on demographic characteristics,

mathematical ability, or recency of exposure to statistical concepts, which rules out several potential confounds that might otherwise explain group differences. With both groups simultaneously enrolled in the introductory statistics course, the results from this study emphasize the value of the new learning technique and the overall influence it may have on statistical literacy for undergraduate students. Effect sizes in the present study were comparable to those reported in previous research (Karazsia, 2013), and they remained large according to Cohen's (1992) standards. Interpretation of visual aids is an important component of quantitative literacy (e.g., APA, 2007), and this learning technique appears to foster student learning in this important competency domain.

Although the present study was the most rigorous thus far to evaluate this teaching technique, one limitation is that group assignment was not random. Random assignment in this ecologically valid setting may be impossible due to student schedule conflicts with other courses. However, we ruled out the likelihood that confounding factors due to the quasi-experimental design explain group differences by exposing both groups to the intervention. Where we exposed both groups to the technique, no group differences emerged. Group differences only emerged when we did not expose the control group to the intervention. Although group differences were in the expected direction, the absolute scores on the assessments were not consistently strong (mean percentages for each group across each time point ranged from 38% to 84%). One interpretation of these percentages is that they indicate poor mean performance on the assessments. However, it is important to note that the assessments were not designed as tests or exams to be graded. The investigators made the assessments very challenging to avoid a potential ceiling effect which would have decreased variance between groups and made it impossible to measure group differences. Therefore, for the purposes of this investigation, the

ecological validity of the assessment was compromised to ensure that a ceiling effect would not occur.

Another limitation is that the instructor was not blind to study hypotheses, though the likelihood that this impacted student learning or study results was minimal for two reasons. First, when we exposed students in the control condition to the learning technique, their achievement did not differ from students in the experimental condition. Second, the instructor was not present during any of the assessment procedures. Despite these limitations, evidence is mounting that the proposed teaching technique fosters student interpretation of tables presented in APA format, which is an important skill for students to learn. As noted by Karazsia (2013), frustration with student inability to interpret tables was the impetus for developing this technique, and there are now two studies documenting this technique's effectiveness. Future research will be important to examine if type of professor feedback influences the effectiveness of the technique. For example, is professor feedback necessary for students to obtain the learning benefits of completing an APA-formatted table, or could students critique their own work by comparing their tables with a template or the work of other students? Exploring these possibilities could enhance the potential applications of this technique to alternative learning environments, such as online courses.

The proposed teaching technique holds promise for courses throughout the psychology curriculum, not just courses in statistics. For example, introductory courses in psychology introduce the correlation coefficient (e.g., Myers, 2009), so implementing this learning technique may enhance student performance and ability within those courses. Additionally, this technique may be useful to students in more advanced psychology undergraduate courses as they consume literature that applies the statistical concepts they learned in previous statistics courses. As noted

previously, with additional research the technique could be extended to self-study learning environments or online courses. Further, this technique may be useful on the graduate level for enhancing and facilitating understanding of advanced statistical procedures and methodologies (such as tables of fit indices from structural equation models).

References

- Acock, A. C., van Dulmen, M., Kurdeck, L., Buehler, C., & Goldscheider, F. (n.d). *Tables for national council on family relations journals*. Retrieved from <http://people.oregonstate.edu/~acock/tables/>
- Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD program in North America. *American Psychologist*, 63, 32–50. doi:10.1037/0003-066X.63.1.32
- Aleven, V., & Koedinger, K. R. (2002). An effective meta-cognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26, 147–179. doi:10.1016/S0364-0213(02)00061-7
- American Psychological Association. (2007). *APA guidelines for the undergraduate psychology major*. Washington, DC: Author. Retrieved from www.apa.org/ed/psymajor_guideline.pdf
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi:10.1037/0033-2909.112.1.155
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294. doi:10.1037/0096-3445.104.3.268
- Dunn, D. S., Brewer, C. L., Cautin, R. L., Gurung, R. A. R., Keith, K. D., McGregor, L., N.,...

- Voiget, M. J. (2010). The undergraduate psychology curriculum: Call for a core. In D. F. Halpern (Ed.), *Undergraduate education in psychology: A blueprint for the future of the discipline* (pp. 47–61), Washington, DC: American Psychological Association.
doi:10.1037/12063-003
- Johnson, M., & Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education*, 14. Retrieved from <http://www.amstat.org/publications/jse/v14n2/johnson.html>
- Karazsia, B. T. (2013). Enhancing table interpretation skills via training in table creation. *Teaching of Psychology*, 40, 208–211. doi:10.1177/0098628313487453
- Karazsia, B. T., Crowther, J. H., & Galioto, R. (2013). Undergraduate men's use of performance- and appearance-enhancing substances: An examination of the gateway hypothesis. *Psychology of Men & Masculinity*, 14, 129–137. doi:10.1037/a0027810
- Karazsia, B. T., Guilfoyle, S. M., & Wildman, B. G. (2012). The mediating role of hyperactivity and inattention on sex differences in paediatric injury risk. *Child: Care, Health, and Development*, 38, 358–365. doi:10.1111/j.1365-2214.2011.01243.x
- Lunsford, M. L., & Poplin, P. (2011). From research to practice: Basic mathematics skills and success in introductory statistics. *Journal of Statistics Education*, 19. Retrieved from <http://www.amstat.org/publications/jse/v19n1/lunsford.pdf>
- Myers, D. G. (2009). *Psychology* (9th ed.). New York, NY: Worth Publishers.
- Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, 5, 297–310. doi:10.1080/19312458.2011.624490

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research* (3rd ed.). New York, NY: McGraw Hill.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Allyn and Bacon.

Table 1

Descriptive Statistics for Demographics and Table Assessments (N = 48)

Variables	Control			Experimental			<i>t</i>
	<i>n</i>	<i>M (SD)</i>	95% CI	<i>n</i>	<i>M (SD)</i>	95% CI	
Age (complete sample)	24	18.76 (0.77)	[18.41, 19.11]	24	18.79 (0.83)	[18.44, 19.14]	0.12
Age (follow-up subsample)	9	18.56 (0.73)	[18.00, 19.11]	15	18.53 (0.74)	[18.12, 18.94]	0.25
Math ability (complete sample)	24	11.62 (2.85)	[10.32, 12.92]	24	10.96 (2.85)	[9.75, 12.16]	0.78
Math ability (follow-up subsample)	9	10.22 (2.91)	[7.10, 12.46]	15	11.47 (2.80)	[9.92, 13.02]	1.04
Dependent Variables							
Time 1: Assessment 1	24	4.14 (1.67)	[3.38, 4.90]	24	5.04 (0.87)	[4.67, 5.41]	2.31*
Time 1: Assessment 2	24	2.88 (1.23)	[2.32, 3.44]	24	2.71 (1.08)	[2.25, 3.17]	0.50
Time 2: Assessment 1	9	2.56 (0.88)	[1.88, 3.23]	15	3.33 (1.11)	[2.72, 3.95]	1.78 ⁺
Time 2: Assessment 2	9	2.28 (1.60)	[1.05, 3.51]	15	2.53 (1.23)	[1.85, 3.22]	0.40

Note. The sample size for the follow-up subsample (Time 2) was reduced due to attrition.

⁺*p* < .10, **p* < .05